

TIRC P. A.
Study
#50

Reprinted from the Proceedings of the NATIONAL ACADEMY OF SCIENCES
Vol. 45, No. 3, pp. 389-393. March, 1959.

SOME STATISTICAL OBSERVATIONS ON A COOPERATIVE STUDY OF HUMAN PULMONARY PATHOLOGY. II

By EDWIN B. WILSON AND MARY H. BURKE

OFFICE OF NAVAL RESEARCH, BOSTON, MASSACHUSETTS, AND TOBACCO INDUSTRY RESEARCH
COMMITTEE, NEW YORK, NEW YORK

Communicated December 26, 1958

In our first paper¹ we gave some general average data for the readings of eight pathologists in eight different cities on slides made from sections taken in standard positions in run-of-the-mill lungs at autopsy, using the following classifications: normal, hyperplasia, metaplasia, atypical metaplasia, carcinoma-in-situ and carcinoma. As carcinoma-in-situ was found so rarely by any of the pathologists, that classification will be combined with atypical metaplasia in this continuation of the study; there will be only five groups and their rank indices² will be 0, 1, 2, 3, 4.

1003537387

When we became convinced that the classification was being made on different bases by the different pathologists, we asked all twelve to read a selected sample of 40 slides. This they kindly did, and we reported on the considerable statistical differences in the readings. As the main object in all the work has been to obtain comparable data in the twelve cities for the degree of pathology in the lungs examined, we stated that it would be well to have a considerable sample of the slides from all cities read by several pathologists. The need for this is clear from the differences shown in Table 1 for the percentages of their slides placed in the 5 groups by the pathologists in eight of the twelve cities.³

TABLE 1

PERCENTAGE DISTRIBUTIONS FOR MALES, AGE 25 AND UP							
Reader	Slides	0	1	2	3	4	Index
J.....	909	28.8	53.6	11.7	4.2	1.8	0.97
D.....	941	57.1	21.1	7.7	11.1	3.0	0.82
A.....	408	38.7	46.1	15.0	0.0	0.2	0.77
E.....	630	66.7	9.7	18.7	3.6	1.3	0.63
B.....	223	65.9	9.4	21.1	2.6	0.9	0.63
L.....	2495	76.4	6.9	11.9	3.3	1.5	0.47
I.....	669	74.4	8.4	16.3	0.9	0.0	0.44
H.....	1418	81.8	9.7	8.0	0.4	0.1	0.27
Mean	...	61.2	20.6	13.8	3.3	1.1	0.62

We were fortunate enough to find three of the pathologists who were willing to read a sample of 609 slides drawn from the different cities by random processes.⁴ We included also the 40 slides previously read by all twelve. The present paper is a report on the results of the rereading. The two sets of slides will be treated separately. The gross results are in Tables 2 and 3.

TABLE 2

DISTRIBUTION OF TOTAL OF 609 SLIDES ON REREADING							
Reader	Slides	0	1	2	3	4	Index
A.....	609	359	93	127	14	16	0.744
E.....	609	348	25	212	6	18	0.885
L.....	609	357	88	133	7	22	0.760
Total	1827	1066	206	472	27	56	0.796

Reader A is high in atypicals (3) and Reader E is low in hyperplasia (1) and high in metaplasia (2) compared with the other two.

TABLE 3

DISTRIBUTION OF THE 40 SLIDES ON REREADING							
Reader	Slides	0	1	2	3	4	Index
A.....	40	4	4	27	3	2	1.875
E.....	40	5	2	28	2	3	1.900
L.....	40	5	6	25	1	3	1.775
Total	120	14	12	80	6	8	1.850

In this small sample, distributed very differently from the large one, the differences noticeable in the latter are not in evidence; but the distribution is significantly different from that previously found by all twelve pathologists, viz., 48, 120, 223, 57, 32; though it is not significantly different from what the three rereaders found, viz., 16, 20, 64, 10, 10.

The rereadings of the 40 slides by the three readers and their original readings

1003537388

different d sample statistical to obtain lungs ex- the slides from the in the 5

have the properties in Table 4. The first reader has not changed his mean significantly, the second has decreased his, and the third increased his, each significantly. The means thus have come closer together. The self-correlation coefficients vary from 0.65 to 0.86.

TABLE 4

Reader	Mean II	Mean I	Mean II - Mean I	Correlation $r_{I, II}$
A.....	1.875	1.800	+0.075 \pm 0.114	0.65
E.....	1.900	2.125	-0.225 \pm 0.103	0.76
L.....	1.775	1.525	+0.250 \pm 0.091	0.86

In the random sample, the numbers of slides belonging to A, E, and L, respectively, were 73, 72, and 60. The comparison of the rereadings by each of his own slides is given in Table 5. It is seen that the three pathologists are reading their own slides about as they did before and that the self-correlation coefficients⁵ are of about the same magnitude as for the 40 slides.

TABLE 5

Reader	Mean II	Mean I	Mean II - Mean I	Correlation $r_{I, II}$
A.....	0.548	0.644	-0.096 \pm 0.089	0.60
E.....	0.792	0.764	+0.028 \pm 0.073	0.81
L.....	1.150	1.333	-0.183 \pm 0.142	0.55

With this background we may turn to the standardization of the percentages over classes which result from using the rereadings of the three pathologists as a basis. The method is similar to that on standardizing death rates for age and sex against the age and sex distributions of a standard population. In Table 1, J put 28.8 per cent of his slides in the normals. The sample drawn for J from his 909 slides and presented to the three pathologists among other slides, contained 32 normals, 59 hyperplasias, 17 metaplasias, 5 atypicals, and 3 carcinomas. These were distributed by the three pathologists (averaged) as given in Table 6. We

TABLE 6

Rank	Number	0	1	2	3	4
0.....	32	31	$\frac{2}{3}$	$\frac{1}{3}$	0	0
1.....	59	$38\frac{1}{3}$	13	$7\frac{1}{3}$	$\frac{1}{3}$	0
2.....	17	3	$\frac{2}{3}$	$10\frac{2}{3}$	$2\frac{2}{3}$	0
3.....	5	2	$\frac{1}{3}$	$1\frac{2}{3}$	0	1
4.....	3	0	0	$\frac{1}{3}$	$\frac{1}{3}$	2

have to assume that all J's slides of each class would have been distributed in these same proportions had they all been reread. Thus his 28.8 per cent of normals in Table 1 would have been distributed as $\frac{31}{32}$ of 28.8 per cent normals, $\frac{1}{48}$ of 28.8 per cent hyperplasia, and $\frac{1}{96}$ of 28.8 per cent metaplasia. In this way one calculates Table 7.

TABLE 7

Original	0 28.8	1 53.6	2 11.7	3 4.2	4 1.8
0.....	27.9	0.6	0.3	0.0	0.0
1.....	34.8	11.8	6.7	0.3	0.0
2.....	2.1	0.4	7.3	1.8	0.0
3.....	1.7	0.3	1.4	0.0	0.8
4.....	0.0	0.0	0.4	0.2	1.2
Standardized	66.6	13.1	16.1	2.3	2.0

1003537389

The comparison of J's original percentages at the top of this table with the adjustment by the averaged readings of the three pathologists reveals the fact that they would have read his slides very differently and would indeed have given for them a percentage distribution not very far from the mean. This does not mean that J was wrong and they are right; it only means that there is a difference. Treating all eight in the same way, Table 1 as adjusted becomes Table 8.

TABLE 8

ADJUSTED PERCENTAGE DISTRIBUTIONS FOR MALES, AGE 25 AND UP							
Reader	Slides	0	1	2	3	4	Index
J.....	909	66.6	13.1	16.1	2.3	2.0	0.60
D.....	941	62.7	13.3	18.0	0.8	5.1	0.72
A.....	408	63.9	15.3	19.9	0.0	0.9	0.59
E.....	630	68.5	6.5	22.5	1.8	0.8	0.58
B.....	223	74.3	7.6	14.0	1.0	3.0	0.51
L.....	2495	60.9	12.2	21.0	0.9	5.0	0.78
I.....	669	71.1	5.2	20.3	0.5	3.0	0.58
H.....	1418	68.8	13.3	13.6	0.6	3.7	0.57
Mean.....	67.1	10.8	18.2	1.0	2.9	0.62

When one compares Tables 1 and 8, bearing in mind that, had any three other pathologists reread the slides, the adjustments would have been different,⁶ and further bearing in mind that the adjustments have been made by scaling up samples in the different cities of from 60 to 75 with one exceptionally large one of 116, it is obvious that most of the differences between the eight cities have disappeared and that it would be very difficult to separate out from the adjusted percentages items which proved that the pathological conditions of the lungs in the different cities were in fact different.⁷

Even when comparisons of general morbidity or mortality conditions in different places are dubious because of differences in reporting, the analysis of local reports by those familiar with local conditions has value. We hope that the individual pathologists who have been good enough to engage in this survey will work up their data in any way they please. We shall be glad if our study furnishes them something of value for theirs.

¹These PROCEEDINGS, 43, 1073-1078, 1957.

²This will make the mean indices, standard deviations, and correlation coefficients of the previous paper not strictly comparable with those here, but the comparison will not have to be made.

³It has been necessary to omit four of the twelve cities. One of the co-operating pathologists failed to send in the data from his city; one had so few cases that it seemed better not to include his city in the rereading; one sent in no slides to be reread; one had used the Swiss roll instead of the standard sections, and we feared this might introduce noncomparability.

⁴The 609 slides are not strictly random because a few more had been drawn randomly, of which some had to be discarded because at least two of the three rereaders felt that they were not good enough to be read at all. It is, however, our belief that this loss did not seriously disturb the randomness.

⁵The self-correlation coefficients have long been used by psychometrists, educational testers, and others to give one estimate of the reproducibility of the data. See, for example, C. Spearman, *The Abilities of Man, Their Nature and Measurement*; T. L. Kelley, *Crossroads in the Mind of Man: A Study of Differentiable Mental Abilities*; J. P. Guilford, *Psychometric Methods*. On pages 411 ff. of the last is given a brief general discussion of various concepts related to reliability. Our index is a rank index, an index of ordinal position. So are many, if not most, of the grades or marks which teachers use. It may be questionable whether one should treat ranks as cardinal numbers, but that is widely done as we are doing it.

The mean value of the three self correlations on the forty slides is 0.76 ± 0.06 , and of those on their own slides is 0.65 ± 0.09 . We have six mutual correlations of the three pathologists in pairs on the forty slides and six on their own; the values of the means are respectively 0.69 ± 0.03 and 0.58 ± 0.04 . Owing to the small numbers in the samples these means have little statistical stability; but so far as the evidence goes, it indicates that the self correlations are not much larger than the mutual correlations. Or in other words, the three pathologists reproduce one another's readings about as well as they reproduce their own—as measured by these correlations. The natural interpretation is that their remaining differences are chiefly fortuitous or random, due to lack of definition and possibly to lack of complete definability of the pathological material. Some slides may be far from clear; should they be discarded? Some may have part of the mucosae lacking; what about them?

If n, h, m, a, c be the fractions (probabilities) of slides of a certain area on which the worst condition is normal, hyperplasia, metaplasia, atypical metaplasia, and carcinoma, what would be the fractions on slides which had twice that area? This question cannot be answered with any information we have; but it is interesting to consider and may suggest interesting research. If the condition revealed by the slide were so widespread that it would occur on both halves of the slide of double area, there would be no differences in the probabilities. At the other extreme where the (worst) condition is so sharply localised that the condition on the slide had no relation to that on an adjacent equal area, the fractions for slides covering a doubled area could be obtained from combinations of terms in the expansion of $(n + h + m + a + c)^2$. As an illustration, if for slides covering a given area, the fractions (probabilities) are $n = .70, h = .10, m = .15, a = .03, c = .02$, then the results for the slides covering twice the area would be 0.49, 0.15, 0.26, 0.06, 0.04, respectively. If the work were to be done over, it might be well to record enough about the conditions appearing on the slides to learn something about their correlations. Such a study might reveal evidence bearing on the question whether in truth the five conditions are in fact successive.

* The two cities, J and H, top and bottom of Table 1, which showed the highest and the lowest values of the index and the lowest and highest percentages of normals, were each first adjusted by using the rereadings of each of the three pathologists, and the results were in fact different, as must be expected; but a study of their similarities indicated that an averaging of the findings of the three pathologists should give not only a stabler result but one which would give a standardization worth having.

† Consider, for example, what the rereading by A, E, L has done to their own previous distributions:

	A Old	A New	E Old	E New	L Old	L New
Index.....	0.77	0.59	0.63	0.58	0.47	0.78
Per cent normal.....	38.7	63.9	66.7	68.5	76.4	60.9

For these three the mean index was 0.62 and has become 0.65—an insignificant change. The differences from the old mean index were $+0.15, +0.01, -0.15$; from the new $-0.06, -0.07, +0.13$. Descriptively the correlation is negative, though not significant. The differences from the respective means of per cent normal were $-21.9, +6.1, +15.8$ and become $-0.5, +4.1, -3.5$, and again the correlation is negative. This is simply an indication of the differences inherent in passing judgments on the slides. If we correlated the two sets of per cent normal in Tables 1 and 8, we would find $r = 0.24$, and if we correlated the two sets of indices, $r = 0.05$. The striking phenomenon to notice is how much the standardization has reduced scatter.

1003537391

1003537392

1003537393

1003537394